

L2-Relaxation

Zhentao Shi and Zhan Gao

April 19, 2022

Forecast combination (Bates and Granger, 1969) is widely used in practical forecasting problems. ℓ_2 -relaxation is an algorithm designed for high-dimensional forecast combinations in the presence of many forecasts. This vignette introduces the R implementation of Shi et al. (2020)'s ℓ_2 -relaxation.

1 Introduction

Let y_{t+1} be an outcome variable of interest, and there are N forecasts, $\mathbf{f}_t := \{f_{it}\}_{i \in [N]}$, available at time t for y_{t+1} , where $t \in [T] := \{1, 2, \dots, T\}$ and $[N] := \{1, 2, \dots, N\}$. We are interested in finding an $N \times 1$ weight vector $\mathbf{w} = (w_1, \dots, w_N)'$ to form a linear combination $\mathbf{w}'\mathbf{f}_t$ to minimize the mean squared forecast error (MSFE) of the estimation error

$$y_{t+1} - \mathbf{w}'\mathbf{f}_t = \mathbf{w}'\mathbf{e}_t,$$

where $\mathbf{e}_t = (e_{1t}, \dots, e_{Nt})'$ with $e_{it} = y_{t+1} - f_{it}$.

Given the forecast error vector and its sample variance-covariance (VC) estimate $\hat{\Sigma} \equiv T^{-1} \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t'$, Bates and Granger (1969) proposed the following constrained minimization problem

$$\min_{\mathbf{w} \in \mathbb{R}^N} \frac{1}{2} \mathbf{w}' \hat{\Sigma} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}' \mathbf{1}_N = 1. \quad (1)$$

where $\mathbf{1}_N$ is an $N \times 1$ vector of ones. Denote the solution to the above constrained optimization problem as $\hat{\mathbf{w}}^{\text{BG}}$. When $\hat{\Sigma}$ is invertible, we can explicitly solve the problem to obtain the optimal solution $\hat{\mathbf{w}}^{\text{BG}} = \left(\mathbf{1}_N' \hat{\Sigma}^{-1} \mathbf{1}_N \right)^{-1} \hat{\Sigma}^{-1} \mathbf{1}_N$. The requirement of the invertibility of $\hat{\Sigma}$ is not guaranteed in high dimensional settings, and in fact $\hat{\Sigma}$ is always singular if $N > T$.

2 ℓ_2 -relaxation

The ℓ_2 -relaxation primal problem is the following constrained quadratic form optimization

$$\min_{(\mathbf{w}, \gamma) \in \mathbb{R}^{N+1}} \frac{1}{2} \|\mathbf{w}\|_2^2 \quad \text{subject to} \quad \mathbf{w}'\mathbf{1}_N = 1 \text{ and } \|\widehat{\Sigma}\mathbf{w} + \gamma\mathbf{1}_N\|_\infty \leq \tau, \quad (2)$$

where τ is a tuning parameter to be specified by the user. Denote the solution to (2) as $\widehat{\mathbf{w}}$.

The following is the CVXR code with the free convex solver ECOS_BB. The function requires only two input arguments: `Sigma` (The sample covariance $\widehat{\Sigma}$) and the tuning parameter `tau` (τ). The default is $\tau = 0$, under which $\widehat{\mathbf{w}} = \widehat{\mathbf{w}}^{\text{BG}}$.

```
rL2_primal <- function(Sigma, tau = 0) {  
  
  N <- nrow(Sigma)  
  w_gamma <- Variable(N + 1)  
  w <- w_gamma[1:N]  
  gamm <- w_gamma[N + 1]  
  
  objective <- Minimize(0.5 * sum_squares(w))  
  constraints <- list(sum(w) == 1,  
    Sigma %*% w + gamm <= tau,  
    -Sigma %*% w - gamm <= tau )  
  
  problem <- Problem(objective, constraints)  
  result <- solve(problem, solver = "ECOS_BB")  
  w_hat <- result$getValue(w_gamma)[1:N]  
  
  return(w_hat)  
}
```

The formulation is very simple. (2) can be written explicitly as

$$\begin{aligned}
& \min_{(\mathbf{w}, \gamma) \in \mathbb{R}^{N+1}} \frac{1}{2} \mathbf{w}' \mathbf{w} \\
& \text{subject to } \mathbf{w}' \mathbf{1}_N = 1 \\
& \quad \hat{\Sigma} \mathbf{w} + \gamma \mathbf{1}_N \leq \tau \mathbf{1}_N, \\
& \quad -(\hat{\Sigma} \mathbf{w} + \gamma \mathbf{1}_N) \leq \tau \mathbf{1}_N.
\end{aligned}$$

The criterion function is quadratic, and there are $2N + 1$ linear constraints.

The following is the `Rmosek` version looks more complicated, but it runs faster if the commercial convex solver Mosek is available.

```

rL2_primal_mosek <- function(Sigma, tau, tol = 1e-7) {
  N <- nrow(Sigma)
  # variable order: w_1, w_2, ..., w_N, gamma, t, s, r

  prob <- list(sense = "min")
  prob$dparam <- list(INTPNT_CO_TOL_REL_GAP = tol)

  prob$c <- c(rep(0, N + 1), 1 / 2, rep(0, 2))

  A_1 <- rbind(
    c(rep(1, N), 0), # sum of weight == 1
    cbind(Sigma, rep(1, N)) # ||Sigma_hat w + gamma||_infty <= tau
  )
  A_2 <- rbind(c(1 / 2, -1, 0), c(1 / 2, 0, -1)) # transformation of the squared l2 norm
  A <- Matrix::bdiag(A_1, A_2)
  prob$A <- as(A, "CsparseMatrix")
  prob$bc <- rbind(
    blc = c(1, -tau * rep(1, N), 1 / 2, -1 / 2),
    buc = c(1, tau * rep(1, N), 1 / 2, -1 / 2)
  )
  prob$bx <- rbind(
    blx = c(rep(-Inf, N + 1), 0, rep(-Inf, 2)),
    bux = rep(Inf, N + 4)
  )
  # conic constraint

```

```

prob$cones <- matrix(list("QUAD", c(N + 4, 1:N, N + 3)))
rownames(prob$cones) <- c("type", "sub")

mosek_out <- Rmosek::mosek(prob, opts = list(verbose = 0))

xx <- mosek_out$sol$itr$xx
w_hat <- xx[1:N]
gamma_hat <- xx[N + 1]
status <- mosek_out$sol$itr$solsta

return(list(
  w = w_hat,
  gamma = gamma_hat,
  status = status
))
}

```

References

- Bates, J. M. and C. W. Granger (1969). The combination of forecasts. *Operational Research Quarterly*, 451–468.
- Shi, Z., L. Su, and T. Xie (2020). High dimensional forecast combinations under latent structures. *arXiv 2010.09477*.